ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

Discovering Complex Trading Rules by Using Genetic Algorithm for Profitable Trading Strategies

K. Yuvana[1], D. Saravanan[2]

Student, Dept. of CSE, Pavendar Bharathidasan College of Engineering & Technology, Trichy , Tamilnadu, India[1]

HOD, Dept. of CSE, Pavendar Bharathidasan College of Engineering & Technology, Trichy, Tamilnadu, India[2]

ABSTRACT: A framework for exploitation of news in stock trading strategies. The strategies take the form of rules that combine technical trading indicators with a news variable, and are revealed to the used for genetic programming. Events are extracted from news messages presented in free text without annotations. Test the introduced framework by deriving trading strategies based on technical indicators and impacts of the extracted events. It is used in share market or financial market. The optimal trading rules are used to find out the news variable, indicating the added value of news for predictive purposes and validating. The proposed work automatically incorporate news in stock trading strategies.

KEYWORDS: Computer applications, evolutionary computing and genetic algorithms, learning, natural language processing, web text analysis**.**

I. INTRODUCTION

Financial markets are driven by information. One important source of information consists of news communicated by different media agencies through a variety of channels. With the increasing number of information sources, resulting in high volumes of text from news, manual processing of the knowledge being communicated becomes a highly difficult task. Additionally, given that this information is time-sensitive, especially in financial markets, selecting and processing all the relevant information in a decision-making process, such as the decision whether to buy, hold, or sell an asset is an especially challenging task. This environment motivates a need for automation in the processing of information, to the extent that investment decisions where the news factor plays an important role can be based on an automatically generated recommendation that takes into account all news messages relevant to a certain asset. Using information extracted from text in a financial context is an area enjoying increasing attention. In (Das and Chen, 2007) the authors extract investor sentiment from stock message boards. The prediction of bankruptcy of firms, as well as fraud, based on textual data from the Management Discussion and Analysis Sections (MD&A) of 10-K reports is investigated in (Cecchini et al., 2010). A popular Wall Street Journal column is used for investigating asset prices as well as trading volumes in (Tetlock, 2007). Financial news stories are used for the prediction of stock returns and firms' future cash flows in (Tetlock et al., 2008). Thus, the qualitative data may emerge from different sources, and can be used for the prediction of different financial aspects of firms' performance. In this chapter we focus on information presented in textual format, taking the shape of news messages relating to some aspect of the financial market, with a particular focus on companies listed under the FTSE350 stock index. The question that we answer is how the information communicated through textual news messages can be automatically incorporated into trading strategies. This requires a two-step approach consisting of: i) extracting the relevant events, as well as the involved entities, from the text of the news messages, and ii) associating an impact with each of the extracted events. Upon extracting the events and associating these with a predefined impact, trading rules based on news can be derived. In this chapter we only consider technical trading indicators as part of these trading rules, but the approach can be easily extended to incorporate other indicators, such as, for example, indicators initiating from fundamental analysis. Technical trading has previously been used with the goal of financial forecasting (Leigh et al., 2002; Mehta and Bhattacharyya, 2004), thus motivating our choice for this approach. The trading strategies that are built are expressed in the form of trees, where the leaves are connected by the logical operators and and or. These trading strategies generate a buy or sell signal

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

for the assets they are applied to, and are determined through genetic programming where a pool of possible trading strategies is tested on historical stock data. We hypothesize that, if the proposed framework is correct, news will be included in the trading strategies generated through genetic programming. Additionally, the trading strategies that we derive in this way should generate positive returns. The first hypothesis comes from the idea that, when providing a genetic program with a pool of variables without the restriction that all these variables should be included in a trading strategy, only the variables that are maximizing the returns will be selected. Trading strategies including a news variable will thus indicate that the content of the news messages has been quantified in such a way that enables generation of profit beyond the ability of trading rules based solely on technical analysis. The second hypothesis states that, next to generating trading strategies based on news, the resulting rules should also be able to obtain a positive return. The chapter is structured as follows. we present previous work related to the relationship between news and the stock market, and the type of events that are proven to influence stock prices. It is provides an initial, quantitative investigation of the relationship between news messages and the stock market, as apparent from the used dataset. The paper is organized as the following sections. Section II describes the related work of Feature based share market or financial market. Section III Existing system. The Experimental results are shown in the Section IV.

II. RELATED WORK

Information drives the markets. Available from an increasing number of sources, most of which are available on-line and to a wide audience, information is more than ever present in everyday life. Different news providers, as well as institutional entities, ranging from stock-listed companies to charities, use the Web as communication channel. While the problem of access to this information is diminishing at a fast pace, new challenges emerge. The most important such challenge consists of selecting just those pieces of information that are relevant to the user and aggregating them for decision-making.

A significant share of the information available today is presented as text. Generally speaking, these texts can contain either facts, opinions, or a combination of both. Those texts containing facts are a collection of (to a higher degree) objective statements about entities orevents, while texts containing opinions are more subjective in nature and convey a sentiment on the subject being described (Liu, 2010). The sources of information themselves are diverse, but here we focus only on those sources deemed relevant for news analytics for financial decision support. They are distinguish between main stream media, press releases of financial entities, (technical) reports of organizations and institutions involved in-, or overseeing, financial markets, message boards and social media, and finally, a source not yet considered in academic efforts in this area, dissident opinions. Main stream media are the main provider of facts and opinions, and the high availability hereof is due mostly to the growth of the World Wide Web.

Large, widely recognized news providers include Associated Press, Reuters, Dow Jones, Google–, and Yahoo! News, and many others, some of them also able to provide annotated news aimed directly at automated trading. Many other news sources populate the on-line landscape, local and international, general or aimed at specific niches in the market, and available in different languages. Financial entities also increasingly use of the Web channel to broadcast information directly to stakeholders and market participants, in the form of press releases, periodical (financial) reports, and news messages regarding different aspects of the business. Companies also increasingly make use of social media for attracting and retaining customers, increasing brand loyalty, but also as communication channel to customers and stakeholders. For example, the official IBM twitter account, has more than 20,000 followers and more than 600 tweets. Boeing Airplanes has over 55,000 followers on Twitter, and has posted more than 3,000 tweets. An example tweet from Boeing announcing an order for the company's main product, posted on June 8th, 2012, is shown below.

Its purpose is to give a basic understanding of how investment decisions are done on a basic level. In the field of finance, investment generally refers to the act of using capital (i.e. money) to buy some type of asset expected to generate a profit for the investor over time. The focus of this report is on stocks, which is a specific type of financial asset, and how stock prices are affected by financial news articles. The theory discussed and developed in the following chapters in this thesis should in theory be possible to apply to many other types of liquid assets (a liquid asset is an asset that has many sellers and buyers). Examples of liquid assets are stocks, bonds1, currencies2, commodities3 and mutual funds4, but this thesis focuses only on stocks.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

A stock market index is a method of measuring the price movements of a collection of stocks.Different indexes measure different collections of stocks using different formulas. Indexes are useful to investors because they give a benchmark by which one can compare the returns of individual stocks. The popular Dow Jones Industrial Average (DJIA) index measures the collective return of 30 large cap (large market capitalization) stocks traded on either the NYSE or the NASDAQ, using a price weighting. This means it is computed by taking a weighted average of the prices of its 30 constituent stocks where each stock is weighted proportionally to its price. we employ genetic programming that can choose be-tween different variables in creating profitable trading rules. The variables originate in technical analysis, except for the news-related variable.

III. EXISTING SYSTEM

Existing system consider minimum number of technical trading indicators.Lexico semantic Patterns for information extraction from news in financial market. The relation between the number of news announcements and trading activity is investigated, the relation between news announcements and monthly returns is also investigated and the relation between earnings announcements and trading volume around the announcement date is investigated. Maximum number of rules cannot be generated in existing system and also day by day information cannot be collected.

IV. PROPOSED SYSTEM

The proposed framework is accurate for including news in technical trading strategies. A genetic program is used to discover complex trading rules based on technical indicators and news based signals. It is used for three step approach..i)Extracting the relevant events from the text of the new message .ii)Associating an impact with each of the extracted events.iii)Making use of the impact of news events in trading strategies. More number of technical indicator rules are used in proposed system and also tofindout the prediction value can be calculated.

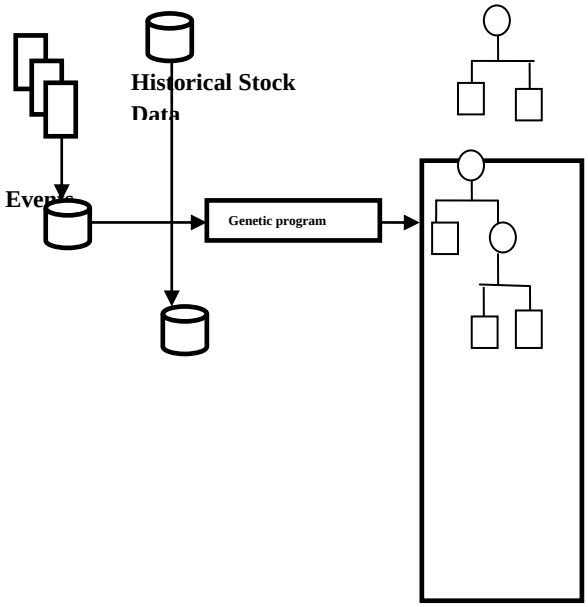

Fig.1 system architecture

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

GENETIC PROGRAMMING:

Genetic programming (GP) is an evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that perform a user-defined task. Essentially GP is a set of instructions and a fitness function to measure how well a computer has performed a task. It is a specialization of genetic algorithms(GA) where each individual is a computer program. It is a machine learning technique used to optimize a population of computer programs according to a fitness landscapedetermined by a program's ability to perform a given computational task.

Algorithm 1 Genetic programming.

Require: α>0: minimum improvement

> 0: maximum times of no improvement > 0: population size
0 < : number of parents

0 1: mutation probability

1: = generateRandomPopulation( )
2: old = *1*, new = calcFitness( ), *b* = 0
3: **while** *b* < **do**
4: addIndividual(′, getBest(, new))

5: **while** *j j* ′ *j* < *j j* **do**

6: = selectRandomParents(, new,)
7: # = crossOver( )

8: #′ = mutate(#,)
9: addIndividual(′, #′)

10: **end while**
11: = ′, old = new, new = calcFitness( )

12: **if** new old **then**

13: *b* = *b* + 1

14: **else if** *b* > 0 **then**

15: *b* = 0

16: **end if**

17: **end while**

18: **return**

TECHNICAL TRADING:

This section focuses on the technical trading indicators used in trading strategies generated via genetic programming. The indicators included in different type of study are: the simple moving average (SMA), the Bollinger band (BB), the exponential moving average (EMA), the rate of change (RoC), momentum (MOM), andmoving average convergence divergence (MACD). The choice for these indicators is based on their widespread use in technical trading .

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

SIMPLE MOVING AVERAGE (SMA),

Simple moving average is the average stock price over a certain period of time. Keep in mind that same weighting is given to each last 20 days of the price.

$$MA_n = \frac{\sum_{i=1}^{n} D_i}{n}$$

where

n = demand of periods in the moving average

D_i = demand in period i

It is many traders watch for short-term averages to cross above longer-term averages to signal the beginning of an uptrend.The buy and sell signal is calculated the last 20 days of the price of a stock

.

EXPONENTIAL MOVING AVERAGE (EMA)

The exponential moving average (EMA) is a product of statistical analysis. I remember listening to my professor talk about EMAs in undergrad and never did I believe I would be using this powerful price indicator to make a living in the financial markets. The EMA is very much like the simple moving average, unless the average is weighted to place emphasis on the most recent price action. The cause many technical analysts prefer the EMA is its ability to reuce the loiter between EMA crosser, which acts as buy and sell trigger for active traders. The EMA is also used to construct some indicators.

EMA= (P*α)+(Previous EMA*(1-α))

P=Current price
α=minimum improvement
N=Number of time periods

In Figure 3, the IDDR curve lies well above the IDR, EDR and IEDR curve for all. This is perfectly acceptable since F-Score values at high levels are more practical. Across all F-Score levels, the largest gap of IDDR over IDR is 0.3. The Proposed IEDR thus achieved a significant improvement over IDR, EDR and IEDR.

In information retrieval, precision is the fraction of retrieved documents that are relevant to search. Similarly, in opinion word or feature exaction algorithm precision is the fraction of retrieved opinion words/ features that are relevant to search.

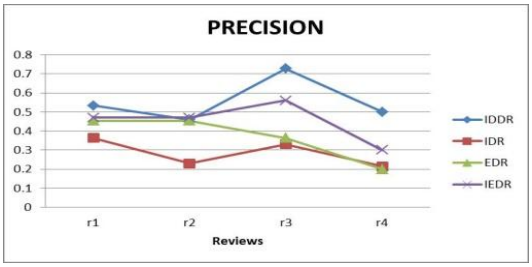

Figure 4: Precision

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

In Figure.4, the IDDR curve lies well above the IDR, EDR and IEDR curve for all. This is perfectly acceptable since Precision values at high levels are more practical. Across all Precision levels, the largest gap of IDDR over IDR is 0.3. The Proposed IEDR thus achieved a significant improvement over IDR, EDR and IEDR.

In information retrieval, recall is the fraction of the documents that are relevant to the query that are successfully retrieved. In opinion word extraction or feature exaction algorithm, recall will be the fraction of the relevant opinion words or features that are relevant to that are successfully retrieved.

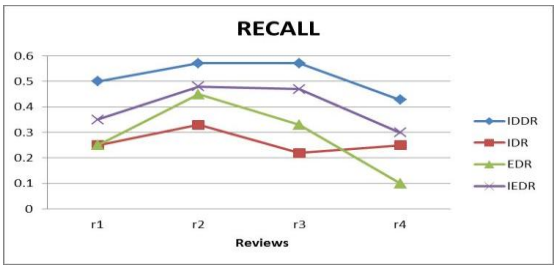

Figure 5: Recall

In Figure 5, the IDDR curve lies well above the IDR, EDR and IEDR curve for all. This is perfectly acceptable since Recall values at high levels are more practical. Across all Recall levels, the largest gap of IDDR over IDR is 0.3 in r3. The Proposed IEDR thus achieved a significant improvement over IDR, EDR and IEDR.

V. CONCLUSION AND FUTURE ENHANCEMENT

Presented a framework for incorporating news into stock trading strategies. The results indicate that the inclusion of news into stock trading strategies can be achieved by extracting the events from the text of the news messages and associating an impact with these events (based on stock price variations for an event). This impact can later be used in the derivation of optimal trading strategies, where the news variable, consisting of impact, is used next to technical indicators. Returning to the two hypotheses stated in the introduction, namely that news will be included in the optimal trading strategies if news is a relevant variable and that these trading rules should generate positive returns, The news variable has been quantified in a meaningful way, confirming first hypothesis that news would be included in the optimal trading strategies. Additionally, all trading strategies that include news events generate positive return, thus confirming second hypothesis.

Future work will focus on including more indicators, technical or non-technical in nature, in the variable pool from which trading strategies are generated. Additionally, a more fine-grained analysis of the news messages, e.g., identification of different event-related information such as the involved actors, should provide multiple information for generating trading strategies. Last, considering the interaction betwixt events occurring within the same day will provide a deeper understanding of the way that news impact stock prices and may lead to more profitable trading strategies.

REFERENCES

[1] J. Borsje, F. Hogenboom, and F. Frasincar, "Semi-automatic financial events discovery based on lexico-semantic patterns,"International Journal of Web Engineering and Technology, vol. 6,no. 2, pp. 115–140, 2010.
[2] W. IJntema, J. Sangers, F. Hogenboom, and F. Frasincar, "A lexico-semantic pattern language for learning ontology in-stances from text," Journal of Web Semantics: Science, Services and Agents on the World Wide Web, vol. 15, no. 1, pp. 37–50,2012.
[3] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web," Management Science,vol. 53, no. 9, pp. 1375–1388, 2007.
[4] M. Cecchini, H. Aytug, G. J. Koehler, and P. Pathak, "Making words work: Using financial text as a predictor of financial events," Decision Support Systems, vol. 50, no. 1, pp. 64–175,2010.
[5] P. C. Tetlock, "Giving content to investor sentiment: The roleof media in the stock market," Journal of Finance, vol. 62, no. 3,pp. 1139–1168, 2007.
[6] ——, "More than words: quantifying language to measure firms' fundamentals," Journal of Finance, vol. 63, no. 3, pp.1437–1467, 2008.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

[7] W. Leigh, R. Purvis, and J. M. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer,neural network, and genetic algorithm: a case study in roman-tic decision support," Decision Support Systems, vol. 32, no. 4,pp. 361–377, 2002.
[8] K. Mehta and S. Bhattacharyya, "Adequacy of training datafor evolutionary mining of trading rules," Decision Support Systems, vol. 37, no. 4, pp. 461–474, 2004.
[9] M.-A. Mittermayer and G. F. Knolmayer, "Text mining systems for market response to news: a survey," Insti-tute of Information Systems University of Bern, Tech. Rep., 2006, berichte/resource/WP-184.
[10] M. L. Mitchell and J. H. Mulherin, "The impact of public information on the stock market," Journal of Finance, vol. 49,no. 3, pp. 923–950, 1994.
[11] W. S. Chan, "Stock price reaction to news and no-news: drift and reversal after headlines," Journal of Financial Economics, vol. 70, no. 2, pp. 223–260, 2003.
[12] S. T. Kim, J. C. Lin, and M. B. Slovin, "Market structure ,informed trading, and analyst' recommendations," Journal of Financial and Quantitative Analysis, vol. 32, no. 4, pp. 507–524,
1997.
[13] S. G. Ewalds, M. B. J. Schauten, and O. W. Steenbeek,"Deinformatiewaarde van kwartaalcijfers," Maandblad voor Ac-countancy en Bedrijfseconomie, no. 7/8, pp. 333–341, 2000.835
[14] J. B. Warner, R. L. Watts, and K. H. Wruck, "Stock prices and top management changes," Journal of Financial Economics, vol. 20, no. 1, pp. 461–492, 1988.
[15] K. A. Bonnier and R. F. Bruner, "An analysis of stock price reaction to management change in distressed firms," Journal of Accounting and Economics, vol. 11, no. 1, pp. 95–106, 1989.06